Sociology 2275, Social Network Analysis, Fall, 2022

Assignment 5: Egocentric Data and Measures

This exercise provides you with a little experience in obtaining egocentric network data from a whole-network data set, in using such data when they are part of a conventional social survey, and working with some common egocentric network measures.

Guidelines in the "Implementation/Software Notes" folder of the course website provide some guidance about egocentric network measures. One set of notes discusses the relevant functions for constructing egocentric network measures from whole network data that are available in xUCINET (note that some of those mentioned in Borgatti et al. do not appear to be working at the present time). Another explains how to work with egoecentric network data in a typical survey.

Preparation: for the first part of this assignment we will work with the data set on "The Corporation" that were studied in assignment 3. Load that into R as you did for assignment 3.

Those data are for a directed network, and it will be convenient to morph them into an undirected one here. This can be done via the `xSymmetrize()` function:

```
xSymmetrize(NET1, Type = "Min")
```

It takes two arguments, one for the network to be studied (as usual) and the other controlling the way in which the data are to be symmetrized: this determines how the edge in the undirected network is produced from the x(i,j) and x(j,i) arcs in the directed network. Options are:

Min—edge value is the minimum value for arcs x(i,j) and x(j,i)
Max—edge value is the maximum value for arcs x(i,j) and x(j,i)
Av—edge value is the average value for arcs x(i,j) and x(j,i)
Sum—edge value is the sum of the values for arcs x(i,j) and x(j,i)
Prod—edge value is the product of the values for arcs x(i,j) and x(j,i) [equivalent to Min for binary data]

This defaults to "Min", but here I suggest that you use "Max". This means, substantively, that a communication edge is deemed to be present if either party mentions it. ("Min" is more restrictive, requiring that both parties acknowledge it.) This is a serious assumption (not just a convenience), and in practice you should consider what it means for data you are working with.

1. **Extract and display the egocentric network for executive 3** (as a data matrix and as a graph).

(This follows Practice 7.6 on the xUCINET website)

To accomplish this, begin by designating executive 3 as "EGO":

```
> EGO <- 3
```

Then find the alters linked directly to the ego, as follows:

```
> Alters_EGO3 <- <Symmetrized Corporation Object>[EGO,]>0
```

Next, be sure that executive 3 is included in the network:

```
> Alters_Ego3[EGO] <- 1
```

This includes the ego node in the list of nodes to be extracted, and also converts vector `Alters_ego3` from a logical to a numeric object.

Then extract the egocentric network (including the ego) from the whole network:

```
> Ego3Network<-<Symmetrized Corporation
Object>[Alters_ego3==1,Alters_ego3==1]
```

You can now display this egocentric network as a matrix and visualize it, as we have done in previous assignments.

2. **Report the egocentric network size and egocentric density for executive 3**.

To do this, it will simplify your life if you remove the ego node from the network. To accomplish this, revisit a couple of the steps in part 1. First, remove the ego node from the alters roster:

```
> Alters_Ego3[EGO] <- 0
```

And then re-extract the egocentric network (this time without the ego node):

```
> Ego3Network<-<Symmetrized Corporation
Object>[Alters_ego3==1,Alters_ego3==1]
```

Then use the `NROW()` and `xDensity()` functions.

3. Based on the logic of Burt's analysis in "Structural Holes and Good Ideas**," which 3 executives are best situated to develop good ideas in "The Corporation," and why? Which 3 executives are least apt to have them?**

For this, use the `xStructuralHoles()` function and the Effective Size and Constraint measures it produces. See software notes on egocentric networks from whole network data. Be sure to use the symmetrized version of the Corporation network when calculating the measures.

4. **How, on average, do the networks of senior and non-senior executives differ in terms of their size, effective size, and constraint?**

There are probably numerous ways of assessing this, but one is to use the `tapply()` function:

```
tapply(<structural hole measures object[,column #],
      <seniority from original (not symmetrizedz) Corporation
      object>,
      mean
      )
```

5. A data set called GSS1985_NetworkExtract.dta including many of the 1985 General Social Survey egocentric network data (together with some respondent-level data) is in the Data folder of the Modules tab of the course's Canvas site. It is in Stata format, but the software note on egocentric network data in surveys explains how to make this useful in R. Begin by doing that.

a. Using the guidance in the software note, construct measures of egocentric network density and the educational composition (average education of alters) for each respondent. Obtain and report descriptive statistics (means and standard deviations) for these measures, as well as egocentric network size (respondent variable "numgiven").

Note: You may leave the education categories (in variables "educ1"-"educ5") as coded in the data set, or recode them to year units by assigning midpoints of the ranges included in each category (see variable labels or GSS website for coding). The latter approach may yield somewhat more interpretable descriptive statistics.

b. Use the egocentric measures you created in part a. together with "numgiven", and the respondent's education (in years, variable "educ"), sex (variable "sex") and age ("age"), to predict variations in general happiness (variable "happy", scored from 1 "very happy" to 3 "not too happy") for the subjects in this study (you may wish to reverse the polarity here by reverse-coding the happiness values; it will make your regression coefficients more interpretable). You may use any reasonable regression method here, including ordinary linear regression/OLS; you may also opt to use some generalized linear model such as ordinal logistic regression.

Describe the results you obtain and what they mean substantively. Does it appear that people with larger networks are happier? Are those who have denser networks more or less happy than those with sparser (less dense) ones? Is happiness associated with educational composition—and if so, how?

Due: Thursday, October 20.

Note: These assignments are not graded, but we do take note of whether or not you do them. Please submit to either derick_baum@g.harvard.edu or pvm@wjh.harvard.edu .